Санкт - Петербургский государственный унивеситет

Кафедра механики управляемого движения

Гриднев Валерий Анатольевич

Выпускная квалификационная работа магистра

Применение метода максимума энтропии в задачах статистики

Направление 01.04.02

Прикладная математика и информатика

научный руководитель,

доктор физ.-мат. наук,

профессор

Шмыров А.С

Санкт-Петербург

2019 г.

Содержание

1 Введение

В наше время трудно переоценить значимость информации, ведь у кого ее больше тот более востребован. Именно поэтому с каждым годом на обработку и добычу информации выделяются огромные средства. Вследствие этого набирают популярность методы и алгоритмы для ее обработки. До 50-х годов прошлого века основное определение энтропии было дано Р.Клазиусом в 1865 году для термодинамических процессов. Однако в 1948 году Клод Шеннон предложил вероятностный подход для актуальной на то время проблемы рациональной передачи информации через зашумленный коммуникационный канал. Его идеи послужили основой для разработки двух основных направлений: теории информации и теории кодирования. Эти области активно развиваются и сейчас, но самое главное, что на основе этих идей он опубликовал две статьи в «Bell System Technical Journal», где и ввел понятия энтропии, как меры случайности. Используя данное понятия энтропии, американский ученый Э.Т. Джейнс сформулировал «принцип максимума энтропии» для решения сложных задач статистики. В настоящее время метод максимума энтропии активно применяют в таких важных областях, как финансы, биометрическая аутентификация, моделирование экстремальных событий. В данной работе будет предложен и показан алгоритм применения метода максимума энтропии к задачам с неполной информацией. При реализации принципа максимума энтропии используется метод множителей Лагранжа, который позволяет перейти от условной оптимизации к безусловной. Данный переход позволяет написать решение задачи оптимизации в параметрическом виде. Результатом работы является алгоритм и программа, разработанная на языке python3, позволяющие получить эффективные оценки параметров бета-распределения, не имея полной информации. Бета-распределение с неполной информацией назовем модифицированным бета-распределением.

2 Проблема оценивания неизвестных параметров распределений случайных величин

Обычно существует два разных подхода для получения точных оценок параметров. А именно классический метод и теоретический подход к решению. Наиболее часто используемые методы при классической оценке заключаются в следующем.

Оценка является одной из основных областей статистического вывода. Статистический вывод - это процесс, с помощью которого выводы из данных выборки используются для получения результатов о населении, из которого была выбрана выборка. Теория оценки была основана профессором Р.А.Фишером в серии фундаментальных работ около 1930 года. Точечная оценка относится к процессу оценки параметра по вероятностному распределению, основанному на наблюдаемых данных из распределения. Это одна из основных тем математической статистики. Проблема оценки, когда некоторый параметр неизвестен, привлекла значительное внимание статистиков в недавнем прошлом. Проблему оценки можно встретить везде: в бизнесе, в науке, а также в повседневной жизни.

Возможно, вы захотите узнать, сколько времени в среднем уйдет на работу, а серьезный садовник может захотеть узнать, какие пропорции некоторых тюльпанов можно ожидать. Если мы рассмотрим такие практические данные, то естественно, что они будут следовать определенному распределению вероятностей некоторой случайной величины. В этом случае мы получаем дистрибутив, и нас могут интересовать его характеристики. Итак, нам необходимо изучить распределение и оценку его параметров, где параметр может быть неизвестен.

Введем основные определения, необходимые для понимания поставленной задачи.

2.1 Основные определения

Определение 1. Случайный эксперимент это эксперимент, результаты которого варьируются и не могут быть предсказаны заранее.

Определение 2. Результат статистического эксперимента называют исходом.

Определение 3. Выборочным пространством статистического эксперимента представляет собой пару (Ω, S), где Ω - пространство элементарных событий (множество всевозможных результатов эксперимента), а S - σ-алгебра событий (система измеримых подмножеств пространства Ω).

Определение 4. Событием называют измеримое подмножество пространства элементарных событий Ω, которое мы рассматриваем. Любое множество $A \in S$ является событием.

Определение 5. Пусть (Ω, S) - выборочное пространство. Нормированая счетно-аддитивная мера P, заданная на S называется вероятностью. Т.е. эта мера удовлетворяет следующим условиям:

$$P(A) \geq 0, \quad \forall A \in S. \tag{1}$$

$$P(\Omega) = 1. \tag{2}$$

Пусть $A_j,\ A_j \in S,\ j = 1, 2, ...$ непересекающиеся множества, т.е $A_j \cap A_k = \emptyset,\ j \neq k$. Тогда

$$P(\cup_{j=1}^{\infty}) = \sum_{j=1}^{\infty} P(A_j). \tag{3}$$

Тройка (Ω, S, P) называется вероятностным пространством.

Определение 6. Пусть (Ω, S) - выборочное пространство. Конечная однозначная функция X, отображающая Ω в R называется случайной величиной, если прообразы всех борелевских множеств из R являются событиями.

Определение 7. Пусть X - случайная величина, определенная на (Ω, S, P). Определим функцию F на R с помощью $F(x) = P\{w : X(w) \leq x\}$ для всех $x \in R$. F неубывающая, $F(-\infty) = 0$, $F(\infty) = 1$. Тогда функция F называется функцией распределения случайной величины X.

В данной работе будут рассмотрены непрерывные распределения и их параметры, поэтому введем необходимые определения.

Определение 8. Пусть X - случайная величина, определенная на (Ω, S, P) с функцией распределения F. Тогда X называется непрерывной случайной величиной, если F непрерывна, то есть если существует неотрицательная функция $f(x)$ такая, что для каждого действительного числа x мы имеем $F(x) = \int_{-\infty}^{x} f(t)dt$. Функция f называется плотностью вероятности случайной величины X.

Если X - абсолютно непрерывная случайная величина, то мы можем определить ее функцию плотности вероятности, как показано ниже.

Определение 9. Любая неотрицательная вещественная функция f может служить функцией плотности вероятности непрерывной случайной величины X, если $f(x) \geq 0$, и удовлетворяет условию

$$\int\limits_{-\infty}^{\infty} f(x)dx = 1. \tag{4}$$

Определение 10. Моменты - это параметры, связанные с распределением случайной величины X. Пусть k - положительное целое число, а c - константа. Если $E[(X-c)^k]$ существует, его называют моментом k-го порядка относительно точки c. Обозначим

$$\mu_k = E(X - E(X))^k. \tag{5}$$

Определение 11. Если $E(X^2)$ существует, дисперсия определяется как $\sigma^2 = var(x) = E(X-\mu)^2$. Величина σ называется стандартным отклонением X. $\sigma^2 = \mu_2 = E(X^2) - (E(X))^2$.

3 Применение принципа максимума энтропии

Распространенная статистическая ситуация касается вывода неизвестного распределения $Q(x)$ из известного распределения $P(y)$, где X (размерность n) и Y (размерность m) имеют известную функциональную связь. Чаще всего $n < m$, и задача относительно проста. Например, если Y_1 и Y_2 являются независимыми случайными величинами, каждая из которых равна $[0, 1]$, можно определить распределение $X = Y_1 + Y_2$; здесь $m = 2$ и $n = 1$. Однако биологические и физические ситуации могут возникать при $n > m$. В общем, при отсутствии дополнительной информации, в этих случаях нет единственного решения Q. Тем не менее, можно все же сделать некоторые выводы о Q. В работе предлагается новый подход максимальной энтропии (MaxEnt), который оценивает $Q(x)$, основываясь только на доступных данных, а именно $P(y)$. Метод имеет дополнительное преимущество в том, что нет необходимости явно вычислять лагранжеву мультипликаторы.

Определение 12. Назовем *энтропией случайной величины* ξ (обозначение H_ξ) интеграл Лебега

$$H_\xi = -\int f_\xi(x) ln f_\xi(x) dx, \tag{6}$$

если этот интеграл существует.

3.1 Формулировка принципа максимальной энтропии

Принцип диктует, что нужно искать распределение, соответствующее доступной информации, которое максимизирует энтропию. Использовать этот принцип можно в силу вариационного свойства, которое формулируется в [1] следующим образом.

Вариационное свойство: экспоненциальное распределение доставляет максимум энтропии на классе абсолютно непрерывных распределений случайных величин, принимающих неотрицательные значения и имеющие фиксированное математическое ожидание.

Теперь рассмотрим метод множителей Лагранжа для нормального и гамма распределений.
Запишем энтропию

$$H_\xi = -\int f_\xi(x) ln f_\xi(x) dx \tag{7}$$

Плотность случайной величины f_ξ должна удовлетворять условиям (1 - условие нормировки, 2,3 - уравнения связи):

$$\int f(x)dx = 1 \tag{8}$$

$$\int x f(x)dx = m \tag{9}$$

$$\int x^2 f(x)dx = m^2 + \sigma^2 \tag{10}$$

Теперь введем множители $\lambda_0, \lambda_1, \lambda_2$ и образуем функцию Лагранжа

$$L = \int (-f(x)lnf(x)) + \lambda_0 f(x) + \lambda_1 x f(x) + \lambda_2 x^2 f(x))dx \tag{11}$$

Продифференцировав подынтегральное выражение по f, получаем для плотности $f^*(x)$ реализующий максимум

$$f^*(x) = -lnf^*(x) - 1 + \lambda_0 + x\lambda_1 + x^2\lambda_2 = 0, \tag{12}$$

откуда

$$f^*(x) = exp(-1 + \lambda_0 + x\lambda_1 + x^2\lambda_2) \tag{13}$$

И теперь, для того чтобы эта формула задавала плотность нормального распределения, нужно, чтобы $\lambda_2 < 0$, тогда (формула) задает нормальную плотность, которая определяется, если заданы дисперсия и математическое ожидание. В итоге получаем

$$f^*(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{14}$$

Теперь рассмотрим гамма-распределение.

Условие нормировки и уравнения связи:

$$\int_0^\infty f(x)dx = 1 \tag{15}$$

$$\int_0^\infty xf(x)dx = m \tag{16}$$

$$\int_0^\infty lnxf(x)dx = l \tag{17}$$

Как и для нормального распределения, зададимся множителями Лагранжа $\lambda_0, \lambda_1, \lambda_2$ и образуем функцию

$$L = \int_0^\infty (-f(x)lnf(x) + \lambda_0 f(x) + \lambda_1 xf(x) + \lambda_2 lnxf(x))dx \tag{18}$$

Подынтегральное выражение является выпуклой вверх функцией от f и имеет единственный максимум при $f = f^*$, который определяем приравнивая к нулю производную по f от подынтегрального выражения

$$-lnf^*(x) - 1 + \lambda_0 + \lambda_1 x + \lambda_2 lnx = 0 \tag{19}$$

откуда

$$f^*(x) = exp(-1 + \lambda_0 + \lambda_1 x + \lambda_2 lnx) \tag{20}$$

4 Оценивание неизвестных параметров модифицированного бета-распределения

Так как бета–распределение является решением задачи о максимуме энтропии (разобрано выше),то и на случай с неполной информацией (бета–распределение на отрезке $[0, 1]$ без интервала $< a, b >$) распространяется принцип максимума энтропии. Сгенерировав выборку и отбросив из нее значения из интервала $< a, b >$, мы задаем параметры. Связи имеют такой же вид, как и при бета-распределении

$$Mln(x) = l_1, \quad Mln(1 - x) = l_2 \tag{21}$$

запишем энтропию

$$H_\beta = -(\int_0^a f_\beta(x) ln f_\beta(x) dx + \int_b^1 f_\beta(x) ln f_\beta(x) dx) \tag{22}$$

Сформулируем вариационную задачу для нашего случая: найти неотрицательную функцию f, определенную на множестве

$M = [0, a] \cup [b, 1]$, удовлетворяющую условиям (1 - условие нормировки, 2,3 - уравнения связи)

$$\int_0^a f_\beta(x)dx + \int_b^1 f_\beta(x)dx = 1 \tag{23}$$

$$\int_0^a ln(x)f_\beta(x)dx + \int_b^1 ln(x)f_\beta(x)dx = l_1 \tag{24}$$

$$\int_0^a ln(1-x)f_\beta(x)dx + \int_b^1 ln(1-x)f_\beta(x)dx = l_2 \tag{25}$$

и доставляющую максимум энтропии.

Воспользовавшись методом множителей Лагранжа, получаем функцию плотности, распределение с таким видом плотности назовем модифицированным бета-распределением.

$$f_\beta = Cx^{\lambda_1}(1-x)^{\lambda_2} \tag{26}$$

В данном случае константа и коэффициенты λ_1, λ_2 неизвестны. Подставим эту функцию в уравнение связи

$$\int_0^a Cx^{\lambda_1}(1-x)^{\lambda_2}dx + \int_b^1 Cx^{\lambda_1}(1-x)^{\lambda_2}dx = 1 \tag{27}$$

$$\int_0^a ln(x)Cx^{\lambda_1}(1-x)^{\lambda_2}dx + \int_b^1 ln(x)Cx^{\lambda_1}(1-x)^{\lambda_2}dx = l_1 \tag{28}$$

$$\int_0^a ln(1-x)Cx^{\lambda_1}(1-x)^{\lambda_2}dx + \int_b^1 ln(1-x)Cx^{\lambda_1}(1-x)^{\lambda_2}dx = l_2 \tag{29}$$

Получена система из трех уравнений с тремя неизвестными, так как не в общем случае интервал $< a, b >$ нам известен, а параметры l_1, l_2 считаются из выборки.

4.1 Решение системы интегральных уравнений

Выразим неизвестную C из уравнения (27)

$$\int_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx = \frac{1}{C} \tag{30}$$

и подставим в уравнения (28 - 29), приходим к системе уравнений относительно двух неизвестных:

$$f_1(\lambda_1, \lambda_2) = \frac{\int\limits_0^a ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx} = l_1 \tag{31}$$

$$f_2(\lambda_1, \lambda_2) = \frac{\int\limits_0^a ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx} = l_2 \tag{32}$$

Нами было реализовано два подхода к решению данной системы, т.к. нет готовых реализаций, позволяющих решать такого рода систему уравнений.

Первый подход: найти решения системы (31 - 32), используя построение сетки на плоскости $l_1,\ l_2$, т.е варьируя значения $l_1,\ l_2$, найти такую пару, которая даст точное равенство для уравнений (31 - 32).

Второй подход: методом градиентного спуска. Посчитаем градиенты:

$$\frac{\partial f_1(\lambda_1,\lambda_2)}{\partial \lambda_1} = \frac{\int\limits_0^a ln^2(x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln^2(x)\lambda_1 x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx} -$$
$$- \frac{(\int\limits_0^a ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx)^2}{(\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx)^2}, \tag{33}$$

$$\frac{\partial f_1(\lambda_1,\lambda_2)}{\partial \lambda_2} =$$
$$= \frac{\int\limits_0^a ln(x)ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(x)ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx} -$$
$$- \frac{\int\limits_0^a ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx}{(\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx)^2} \times$$
$$\times (\int\limits_0^a ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx + \int\limits_b^1 ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx), \tag{34}$$

$$\frac{\partial f_2(\lambda_1,\lambda_2)}{\partial\lambda_1}=$$

$$=\frac{\int\limits_0^a ln(x)ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 ln(x)ln(1-x)\lambda_1 x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx}-$$

$$-\frac{\int\limits_0^a ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx}{(\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx)^2}\times$$

$$\times(\int\limits_0^a ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 ln(x)x^{\lambda_1}(1-x)^{\lambda_2}dx), \tag{35}$$

$$\frac{\partial f_2(\lambda_1,\lambda_2)}{\partial\lambda_2}=\frac{\int\limits_0^a ln^2(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 ln^2(1-x)\lambda_1 x^{\lambda_1}(1-x)^{\lambda_2}dx}{\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx}-$$

$$-\frac{(\int\limits_0^a ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 ln(1-x)x^{\lambda_1}(1-x)^{\lambda_2}dx)^2}{(\int\limits_0^a x^{\lambda_1}(1-x)^{\lambda_2}dx+\int\limits_b^1 x^{\lambda_1}(1-x)^{\lambda_2}dx)^2}. \tag{36}$$

Используя градиентный метод и формулы (33 – 36), найдем значения $\lambda_1,\ \lambda_2$, удовлетворяющие системе (31, 32). Подставив полученные значения $\lambda_1,\ \lambda_2$ в уравнение (30), получим неизвестную C, а следовательно и решение системы (27 – 29).

4.2 Оценивание параметров

Стоит также заметить, что стандартное бета–распределение и полученное модифицированное бета–распределение входят в семейство экспоненциальных распределений. Обратимся к [2] и вспомним определение экспоненциального семейства.

Пусть $\theta = \theta_1, \ldots, \theta_k$ $k-$мерный параметр и плотность $f_\theta(x)$ представляется в виде

$$f_\theta(x) = h(x)exp(\sum_{j=1}^{k} \alpha_j(\theta)U_j(x) + V(\theta)) \tag{37}$$

$$f_\theta(x) = h(x) \tag{38}$$

где все функции, входящие в правую часть, конечны и измеримы.

Так как наше распределение входит в семейство экспоненциальных распределений, то, следовательно, мы можем получить эффективные оценки наших параметров. Оценивать мы будем с помощью статистик типа выборочного среднего.

$$l_1^* = \frac{1}{n}\sum_{k=1}^{n} ln(x_k) \tag{39}$$

$$l_2^* = \frac{1}{n}\sum_{k=1}^{n} ln(1 - x_k) \tag{40}$$

Оценка математических ожиданий выборочным средним является эффективной и несмещенной оценкой. Это следует из неравенства Рао – Крамера.

В приложении представлена программа на языке Python, реализующая решение системы (27 – 29).

5 Результаты и выводы

В результате работы было показано, что метод основанный на принципе максимума энтропии допускает довольно простую алгоритмизацию. Эта алгоритмизация была реализована на примере выборки специального вида, компоненты которого имеют модифицированное бета–распределение. Для оценивания параметров связи, определенных уравнениями связи были предложены эффективные оценки. Оценки получены в явном виде через элементы выборки. Для определения параметров, описывающих плотность модифицированного бета–распределения, была выведена система интегральных уравнений (27 - 29). С помощью двух программ написанных на языке python3, неизвестные параметры были найдены двумя способами. Программа, реализующая сетку, подходит для более частных задач (выборка менее 10^6). Программа, реализующая градиентный метод, подходит для большего количества задач, но в данном случае остается открытым вопрос локальных экстремумов. Обе программы представлены в приложении.

6 Заключение

В данной работе проведено исследование о возможности оценивания параметров распределения с неполной информации. В результате работы предложен алгоритм получения такого распределения на основе принципа максимума энтропии с использованием метода множителей Лагранжа. Получены эффективные оценки параметров модифицированного бета-распределения. Решена двумя способами система интегральных уравнений, связывающих параметры плотности модифицированного бета–распределения и эмпирические моменты. Применение методов сетки и градиентного проиллюстрировано на численом примере.

6.1 Численный пример

Смоделируем бета-распределение с параметрами $\lambda_1 = 2,\ \lambda_2 = 3$.

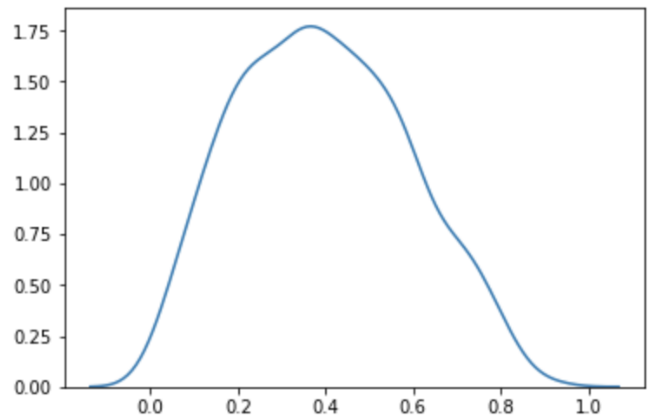

Рис 1. Бета-распределение $(\lambda_1 = 2,\ \lambda_2 = 3)$

При $\lambda_1 = 2,\ \lambda_2 = 3 \Rightarrow l_1 = -1.171$ и $l_2 = -0.569$. Удалим информацию на промежутке $[0.4,\ 0.6]$. Найдем $l_1,\ l_2$ подставим их в систему (31 – 32) и применив методы описанные ранее, получаем значения $\lambda_1,\ \lambda_2,\ l_1,\ l_2$:

Сетка: $\lambda_1^{(1)} = 1.05,\ \lambda_2^{(1)} = 2.05,\ l_1^{(1)} = -1.170,\ l_2^{(1)} = -0.573$;

Градиентный метод: $\lambda_1^{(2)} = 1.02,\ \lambda_2^{(2)} = 2.04,\ l_1^{(2)} = -1.172,\ l_2^{(2)} = -0.571$;

По результатам видно, что точность до второго знака включительно. Таким образом, мы нашли неизвестные параметры модифицированного бета-распределения двумя способами.

7 Список литературы

1. Шмыров А.С., Шмыров В.А. Теория вероятностей: учебное пособие.2012. стр. 162-180.

2. Боровков А.А. Математическая статистика: Учебник.4-е изд. 2010. Стр.178-200.

3. E.T. Jaynes. Probably theory: The logic of science. 1995. Chapter 11.

4. Боровков А.А. Теория вероятностей. 3-е издание. 1999.

5. Гриднев В.А. Оценивание параметров модифицированного бета-распределения: Вкр. 2017

6. Sulagna Mohanty. Estimation of Parameters of Some Continuous Distribution Functions: Master of Science. 2012

7. Roger Levy. Probabilistic Models in the Study of Language. Chapter 2. 2012

8. Jaynes ET. Information Theory and Statistical Mechanics. Phys Rev. 1957;106(4):620-30.

9. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2006.

8 Приложение

В этом разделе, представлен код, реализующий решение системы (27-29), на языке Python3.

8.1 Метод сетки

```
import random

my_list = []
for i in range(1000):
    my_list.append(random.betavariate(2, 3))



a = 0.05
b = 0.95
def cut_out(any):
    new_list = []

    for i in any:
        if (i < a) or (i > b):
            new_list.append(i)

    return new_list
```

```
our_list = my_list.copy()
our_list = cut_out(my_list)
```

```
import numpy as np
import matplotlib.pyplot as plt

_, bins, _ = plt.hist(my_list, bins=50, density=True)
_ = plt.hist(our_list, bins=bins, density=True)
plt.show()
```

```
import seaborn as sns

sns.distplot(my_list, hist=False);
```

```
import matplotlib.pyplot as plt
import numpy as np

gaussian_numbers = np.random.randn(1000)
plt.hist(np.array(my_list))
```

```
plt.hist(np.array(our_list))
```

```
import math

def our_log_l1(x): return math.log(x)
def our_log_l2(x): return math.log(1 - x)

l1 = sum(list(map(our_log_l1, our_list)))/len(our_list)
print(l1)
l2 = sum(list(map(our_log_l2, our_list)))/len(our_list)
print(l2)
```

```
import numpy as np
import scipy.integrate as integrate

def f_1(x):
    return x**lambda1*(1 - x)**lambda2

def f_2(x):
    return math.log(x)*x**lambda1*(1 - x)**lambda2
```

```
def f_3(x):
    return math.log(1 - x)*x**lambda1*(1 - x)**lambda2

znak = 3
shag = 0.05

for i in np.arange(0.0, 10.5, shag):
    for j in np.arange(0.0, 10.5, shag):
        lambda1, lambda2 = i, j

        l1_test = (integrate.quad(f_2, 0, a)[0] +
        integrate.quad(f_2, b, 1)[0])/(integrate.quad(f_1, 0, a)[0]
        + integrate.quad(f_1, b, 1)[0])

        if round(l1_test, znak) == round(l1, znak):

            l2_test = (integrate.quad(f_3, 0, a)[0] +
            integrate.quad(f_3, b, 1)[0])/
            (integrate.quad(f_1, 0, a)[0]
            + integrate.quad(f_1, b, 1)[0])

            if round(l2_test, znak) == round(l2, znak):
                c_test = 1/((integrate.quad(f_1, 0, a)[0] +
```

```
integrate.quad(f_1 , b, 1)[0]))
c = round(c_test, znak)
l1_znac = ((integrate.quad(f_2, 0, a)[0] +
integrate.quad(f_2, b, 1)[0]))*c
l2_znac = ((integrate.quad(f_3, 0, a)[0] +
integrate.quad(f_3, b, 1)[0]))*c
print(i,j,c)
```

8.2 Градиентный метод

```
import math
import numpy as np
import scipy.integrate as integrate
from scipy.optimize import minimize

def f1(x):
    return x**y[0]*(1 - x)**y[1]

def f2(x):
    return math.log(x)*x**y[0]*(1 - x)**y[1]

def dif_f1_lambda1(x):
    return math.log(x)*x**y[0]*(1 - x)**y[1]

def dif_f1_lambda2(x):
    return math.log(1 - x)*x**y[0]*(1 - x)**y[1]

def dif_f2_lambda1(x):
    return math.log(x)**2*x**y[0]*(1 - x)**y[1]

def dif_f2_lambda2(x):
```

```
    return math.log(x)* math.log(1 - x)*x**y[0]*(1 - x)**y[1]

def denominator(y):
    return integrate.quad(f1, 0, a)[0] * integrate.quad(f1, b, 1)[0]

def dif_denominator_lambda1(y):
    return integrate.quad(dif_f1_lambda1, 0, a)[0] * integrate.quad(dif_

def dif_denominator_lambda2(y):
    return integrate.quad(dif_f1_lambda2, 0, a)[0] * integrate.quad(dif_

def l1_test(y):
    return (integrate.quad(f2, 0, a)[0] + integrate.quad(f2, b, 1)[0])/(


def gradient_method(y):
    gradient = [0, 0]

    first_lambda1 = (integrate.quad(dif_f2_lambda1, 0, a)[0] * integrate
    second_lambda1 = (integrate.quad(f2, 0, a)[0] * integrate.quad(f2, b
    gradient[0] = first_lambda1/denominator(y) - first_lambda1/denominat

    first_lambda2 = (integrate.quad(dif_f2_lambda2, 0, a)[0] * integrate
```

```
    second_lambda2 = (integrate.quad(f2, 0, a)[0] * integrate.quad(f2, b
    gradient[1] = first_lambda2/denominator(y) - first_lambda2/denominat

    return gradient

a, b = 0.3, 0.6

for i in np.arange(0.0, 2.0, 0.1):
    y = [0.1 + i, 0.2 + i]
    res = minimize(l1_test, y, method='BFGS', jac=gradient_method, optic
```